DOCUMENT RESUME

ED 254 928 EA 017 592

AUTHOR Kimball, George H.; And Others
TITLE Aptitude-Treatment Interactions in Student Achievement: Implications for Program Policy Decisions.
PUB DATE Mar 84
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 23-27, 1984).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Aptitude; *Achievement Gains; Achievement Tests; Analysis of Covariance; Economically Disadvantaged; Educationally Disadvantaged; *Educational Status Comparison; Elementary Secondary Education; *Federal Legislation; *Federal Programs; Mathematics Achievement; Pretests Posttests; *Program Effectiveness; Reading Achievement; School Statistics; Test Results
IDENTIFIERS California Achievement Tests; *Elementary Secondary Education Act Title I; Oklahoma City Public Schools

ABSTRACT

This study was conducted to determine whether the Chapter 1 instruction provided to low-achieving students in the Oklahoma City Public Schools had a detectable effect on student academic growth. Reading and math achievement test scores of Chapter 1 students in grades 2 to 8 were compared with those of a matched group of low-achieving non-Chapter 1 students over 2 years (1981-82 and 1982-83). Chapter 1 students were provided with 30-50 minutes per day of extra remedial instruction in reading or math, and the achievement of both groups was measured by the California Achievement Test (CAT). An Analysis of Covariance (ANCOVA) procedure was used to address aptitude-treatment interaction (ATI) questions of whether Chapter 1 treatment affected students differently, depending on their pretreatment achievement standing. Results, portrayed by a series of graphs and tables, indicate that for nearly all grades, 23 out of 28 within-grade comparisons showed slope differences in the same direction. These results indicate that initially lower achievers benefit more from Chapter 1 treatment than initially higher achievers. Implications of these findings for Chapter 1 selection criteria are discussed. (TE)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED254928

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

George H. Kimball

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

# APTITUDE-TREATMENT INTERACTIONS IN STUDENT ACHIEVEMENT: IMPLICATIONS FOR PROGRAM POLICY DECISIONS

George H. Kimball, John Crawford,

and Allison Patrick

PLANNING, RESEARCH, AND EVALUATION DEPARTMENT
OKLAHOMA CITY PUBLIC SCHOOLS

Running head: Aptitude-Treatment Interactions

March, 1984

Prepared for presentation at the American Educational Research Association annual meeting, New Orleans, April, 1984.

EA 017 592

# Aptitude-Treatment Interactions in Student Achievement: Implications for Program Policy Decisions

Evaluators are faced with choices among many different methods of analyzing compensatory education (e.g., Chapter I) students' academic growth. The most common method is the Model A (Tallmadge and Wood, 1976) "raw" NCE (Normal Curve Equivalent) gains analysis. Those results address the comparison of Chapter I students' progress in relation to a national norming sample. Work by Kimball and Crawford (1983) and others (Murray and Arter, 1980; Campbell and Stanley, 1966) has demonstrated the problems in using such "underspecified" evaluation models for inferential purposes. In fact, the results from Model A analyses are best suited for descriptive purposes only (e.g., the mean NCE gain does tell us something about the performance level of the Chapter I students as a whole, but statements about the effectiveness of the program or any "treatment effect" cannot be based on simple NCE gain analyses). For a more complete description of the problems and pitfalls concerning the use of Model A analyses, see Murray and Arter (1980).

Our approach has been to attempt to base any inferences about Chapter I treatment effects on between-group studies. It is true that any quasi-experimental design employed will yield less-than-

perfect internal validity, but this is still preferable to the use of treatment-only raw gains analyses for inferential purposes.

Of the three models adopted from the Tallmadge and Wood (1976) effort, the between-group model is most clearly represented as Model B. Since random assignment of students into treatment and comparison groups is not possible, the design is, by necessity, quasi-experimental. Echternacht (1980) noted that the Model B design "is rarely used because its use requires withholding Title I services from students who might nominally be expected to receive some compensatory program" (p.6). However, in our district, it is possible to create "matched" groups by using similar, low achieving, low SES (socioeconomic status) students who are not receiving Chapter I as the "status quo" or comparison group in inferential analyses. The procedures employed in creating the matched groups is described in the method section.

It should be stated that although we do attend to regression lines and utilize regression analyses, we are not following the procedures of Model C (the special regression model) which calculates Title I students' variation from a regression line calculated on higher scoring students in the comparison group.

As part of our application of Model B procedures, an Analysis of Covariance Model was used. This was done for several reasons. Even with the matching procedure on prescores and SES, which causes

the covariance-adjusted means to be identical to the unadjusted (raw) means, the ANCOVA procedure gives more "precise" estimates of effects (lower "Error" sums of squares). This is true whenever the pre-score is correlated significantly with post-score. And, perhaps more importantly, the analyses required prior to ANCOVA address important substantive questions --namely, the "aptitude-treatment interaction" (ATI) questions. The homogeneity of slopes test required prior to ANCOVA poses the question "Did the treatment affect students differently, depending on their pre-treatment achievement standing?"

Even six years after the publication of the handbook of ATIs (Cronbach and Snow, 1977), it is not universally recognized by practicing evaluators that the classic "homogeneity of slopes" test is also the test of ATIs. If the post-on-pre regression lines for the groups being compared are not parallel, and if those lines intersect, and if the point of intersection corresponds to a reasonable pre-score (e.g., within the range of actual data), then there would be evidence that the program primarily benefitted students below or above the pre-score that corresponds to the point of intersection. Such a result would have obvious implications for inferences about program effects, as well as implications for policy concerning student eligibility to receive the services.

What theoretical foundation exists for the expectation that a compensatory education program such as Chapter I will show detectable effects on student achievement scores? There is the work by John Carroll from two decades ago (Carroll, 1963) as well as recent empirical evidence from National Institute of Education (Denham and Lieberman, 1980) labs and centers that indicates that the amount of student engagement with academic content (or, "academic learning time") can be a significant predictor of how much the students learn. Therefore, if the time spent in Chapter I instruction is "extra" time spent on remediation of weaknesses in basic skills, one is directly led to the prediction that program students' performance in basic skills should be greater than what it would be without the Chapter I instruction. Recent research in this district, (Crawford, Patrick, and Kimball, 1984) has also shown that significant (though weak) positive relationships do exist between the amount of time students spent in Chapter I labs and their subsequent achievement gains. However, neither raw NCE gains analyses, nor relationships between allocated and engaged time and achievement gains have addressed the question: How do Chapter I students' achievement gains compare with a local, "matched" sample of similar, low achieving, low SES, non-Chapter I treated students?

The purpose of this study was to determine whether the Chapter I

instruction provided to some of the low achieving students in the district appears to result in a detectable effect on student academic growth. The Chapter I students were compared with a similar ("matched") group of non-Chapter I students. The non-Chapter I students are also low achieving, but do not receive Chapter I services because they either go to school at sites which did not qualify as Chapter I schools, or were not served by Chapter I within their school. Qualification for free or reduced lunches is dependent on family income. Therefore, Chapter I and non-Chapter I sites differ in the average SES (socioeconomic status) of the attending students. Even if Chapter I and non-Chapter I comparison groups were selected that match on initial (pre-treatment) achievement, they would very likely still differ on SES. For this reason, as described in the "Method" section that follows, students in the treatment and comparison groups were matched on both pre-achievement measures and SES as measured by eligibility for free or reduced-payment lunches.

## Method

Subjects. All students entering into these analyses (spanning two school years, 1981-82 and 1982-83) were enrolled in the district for the entire school year. For the analyses of the '81-'82 year, each student had achievement scores from the

district-wide administration of the California Achievement Test (CAT) in May 1981 and May 1982. For the analyses of the '82-'83 year, each student had May 1982 and May 1983 CAT scores. All were in grades 2-8. Both years' groups consisted of both Chapter I students and their matched or "yoked" non-Chapter I counterparts.

Matching Procedure. As mentioned previously, to obtain a more precise measure of the Chapter I treatment effect, the students participating in Chapter I (district-wide) were "matched" with equivalent low-achieving students who were not served by Chapter I. It should be reemphasized that Chapter I school sites are selected based on the percentage of student eligible for free or reduced-payment lunches. Within the site, participation in Chapter I is based strictly on achievement scores. Therefore, in each Chapter I site there are low-achieving students who are eligible for free or reduced-payment lunches as well as those whose family income is above the cutoff. Therefore, this matching process was designed to produce Chapter I and non-Chapter I comparison groups that were perfectly matched on prescores (means and variances) and SES. The Chapter I population was utilized as the "reference" group, and prescores and free lunch eligibility were compared with individuals in the non-Chapter I group, student-by-student (within each grade), to build a sample of "yoked" Chapter I and non-Chapter I groups

who therefore have precisely the same N's, the same means, and the same variances on the prescore measures (and the same N of students receiving free and reduced-payment lunches). This matching process was accomplished separately for reading and math achievement scores, and resulted in comparison groups (Chapter I and non-Chapter I together) of 3,556 (Reading 1982); 3,150 (Math 1982); 3,560 (Reading 1983); 3,134 (Math 1983).

Procedure. The Chapter I "treatment" was provided by the district's Chapter I staff during the 1981-82 and 1982-83 academic years. Elementary students were pulled from their regular classroom for 30-50 minutes per day of reading instruction and/or 30 minutes per day for math. In the elementary grades, the activity that the students are pulled from to go to the Chapter I lab is typically not reading or math activity. Therefore, the time spent in Chapter I labs may be considered "extra" or supplementary time in instruction in the basic skills. In middle school grades, the situation varies somewhat between schools, although students were usually pulled from language arts classes for reading and from math classes for Chapter I math instruction. Even so, the time spent in Chapter I is intended to be more focused on remediation in basic skills than time spent in the regular classroom.

The nature of the reading and math instruction in Chapter I is also qualitatively different from regular classroom instruction. The number of students with each Chapter I teacher is limited to 10 in reading, and 12 in math. Paraprofessional aides are also employed in each Chapter I Learning Center, and the goal is to attend to students' individual needs through one-to-one interactions and discussion between the teacher and individual students. Recent data (Crawford, 1983) collected in a "process" evaluation (using objective classroom observation data) indicated that time spent in Learning Centers is largely academically-oriented and that there are relatively high rates of private one-to-one interactions concerning basic skills content.

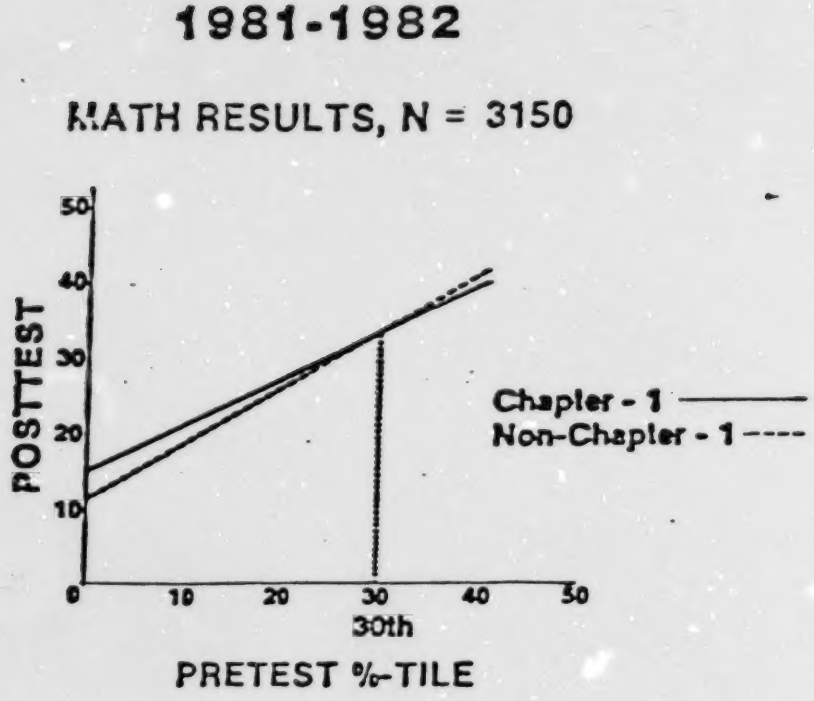
Instruments. The achievement data analyzed as pre and post-scores come from the California Achievement Test (CAT) published by CTB/McGraw-Hill, Monterrey, California. The prescores were derived from either the May, 1981, or May 1982, district-wide testing. The posttest (dependent variable) data came from the May, 1982, or May, 1983, testing with the CAT. The analyses were carried out for the 1981-82 and for the 1982-83 academic years (for 1981-82, pre-data were May, 1981 scores and for 1982-83, pre-data were May, 1982 scores). The scores that were used for analyses purposes were total math and total reading normal curve equivalent (NCE) scores.

NCE scores are conceptually similar to national percentiles, but have the advantage of being equal interval. Spring-to-Spring data were examined since the non-Chapter I students were not tested in the fall. In the figures to follow, NCE scores were converted to national percentiles for convenience of interpretation.

Results

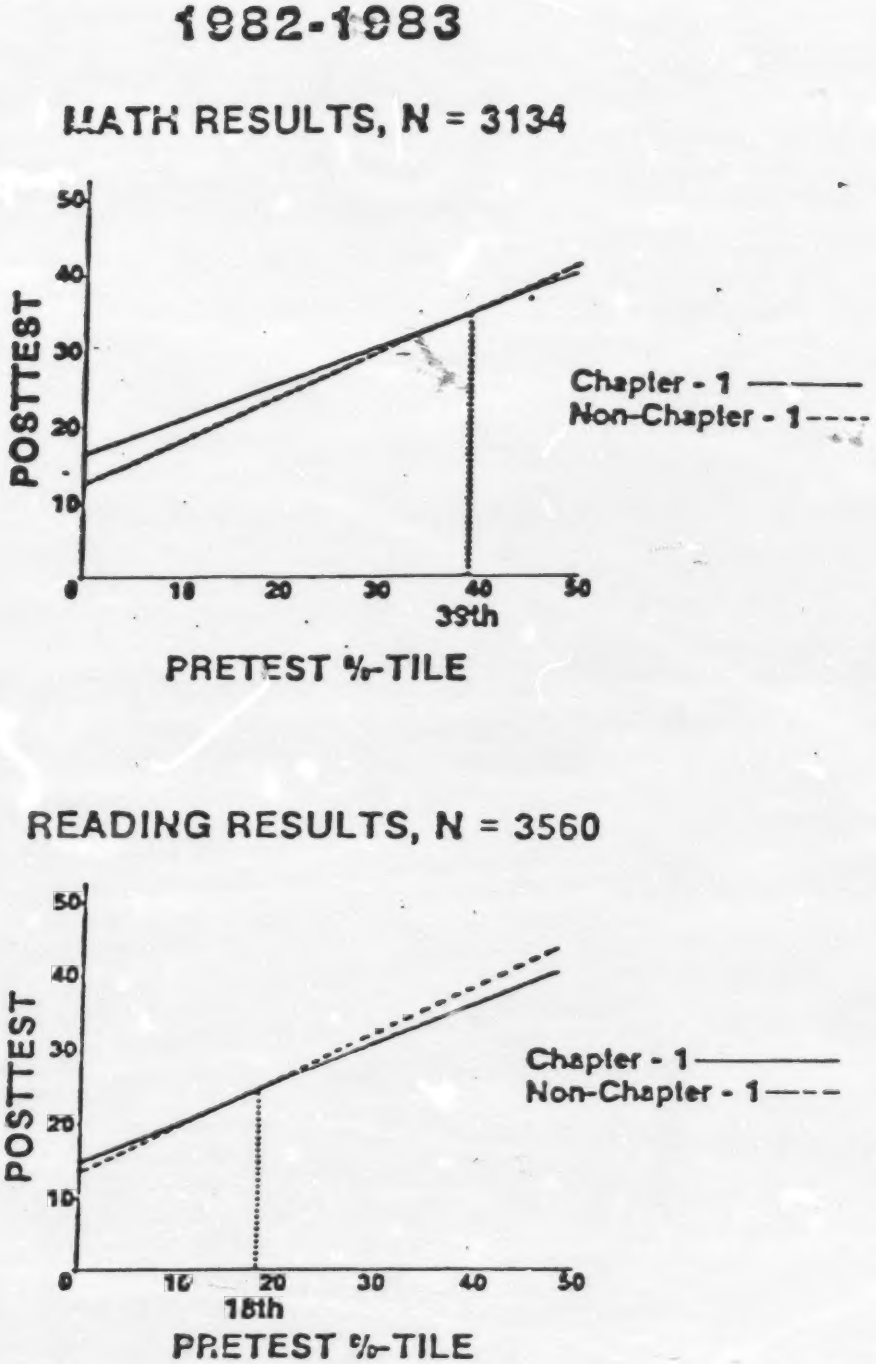
In this study, the homogeneity of slopes test revealed a similar pattern of ATI for both years (1981-82 and 1982-83) and both subject areas (reading and math). The result was consistent across nearly all grades, as 23 out of 28 within-grade comparisons showed slope differences in the same direction as the overall result. Figure 1 shows the overall result (ignoring grade) for math and reading in 1981-82, and Figure 2 presents the results for the 1982-83 analysis.

FIGURE 1




BEST COPY

FIGURE 2




BEST COPY

As Figures 1 and 2 show, the students who benefitted the most from Chapter I "treatment" were those who were initially the lowest achievers. The pattern of ATI's generally replicated across years and across grades. In 23 of the 28 within-grade comparisons the Chapter I students had a lower slope and higher intercept than the non-Chapter I students. Table 1 gives the within-grade comparisons of slopes for both years and both subject matters.

TABLE 1

Slope for Chapter I (CI) and Non-Chapter I (NCI)
Groups by Grade and by Year

| MATH | | | | | READING | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Grade | 1981-82 | | 1982-83 | | Grade | 1981-82 | | 1982-83 | |
| | NCI | CI | NCI | CI | | NCI | CI | NCI | CI |
| 2 | .66 | .64 | .44 | .29 | 2 | .58 | .50 | .42 | .33 |
| 3 | .64 | .53 | .66 | .49 | 3 | .59 | .57 | .70 | .57 |
| 4 | .74 | .68 | .86 | .73 | 4 | .92 | .79 | .81 | .78 |
| 5 | .88 | .78 | .73 | .73* | 5 | .87 | .62 | .68 | .62 |
| 6 | .76 | .59 | .85 | .61 | 6 | .62 | .59 | .68 | .61 |
| 7 | .69 | .71* | .57 | .44 | 7 | .51 | .63* | .60 | .67* |
| 8 | .78 | .43 | .50 | .69* | 8 | .73 | .60 | .42 | .37 |
| Overall Slope | .74 | .64 | .69 | .59 | Overall Slope | .69 | .61 | .63 | .58 |
| Inter-cept | 11.15 | 15.02 | 12.12 | 16.45 | Inter-cept | 11.88 | 14.10 | 13.32 | 14.89 |

*indicates grades with slope differences in opposite direction from the overall slope direction.

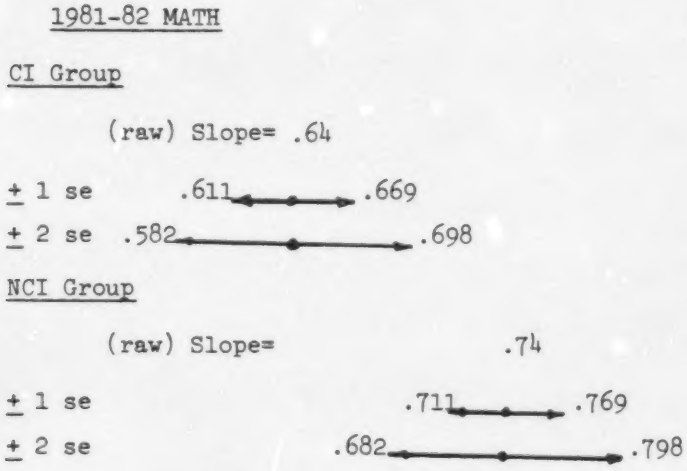
As indicated, there was a great deal of consistency across the grades in the direction of slope differences between the treatment (CI) groups and the comparison groups (NCI), with the exception of grade 7. Why grade 7 differs in 3 out of the 4 comparisons is not known. Nevertheless, there is no a priori reason to

expect this degree of consistency (in 23 of 28 comparisons) in a complex student population such as this, spanning so many grades, and for two academic years.

We have not presented tests of statistical significance of the slope differences using the F-distribution, nor have we employed "regions of significance" methods (Johnson and Neyman, 1936) to analyze the intersection of regression lines, since it may be argued that these data are basically "population" data. The students entering into the analyses are all of the Chapter I students for whom a match could be found (plus their matching, non-Chapter I counterparts). The reader may wonder about the size of the reported slope differences in comparison to the standard errors of the slopes, therefore the last figures of results were prepared (see Figures 3 and 4). For the overall analyses, we created confidence intervals around each observed slope; one at the $\pm$ 1 SE level and one at the $\pm$ 2 SE level. In three of the four comparisons, the Chapter I and non-Chapter I slopes did not overlap with $\pm$ 1 SE around the observed slopes, but did (barely) overlap when the $\pm$ 2 SE confidence interval was used. Only reading scores for '82-'83 showed overlap with $\pm$ 1 SE. There findings tend to support the consideration that the between-group slopes are different.

Figure 3

| 1981-82 MATH | 1981-82 READING |
|---|---|
| CI Group<br><br>(raw) Slope= .64<br>Intercept= 15.02<br>se Slope= .029 | CI Group<br><br>(raw) Slope= .61<br>Intercept= 14.1<br>se Slope= .025 |
| NCI Group<br><br>(raw) Slope= .74<br>Intercept= 11.15<br>se Slope= .029 | NCI Group<br><br>(raw) Slope= .69<br>Intercept= 11.88<br>se Slope= .025 |



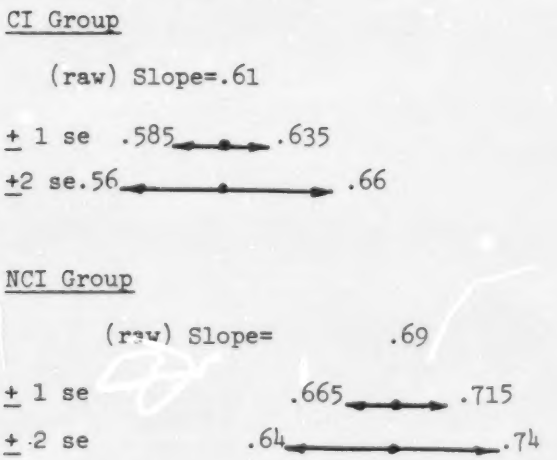
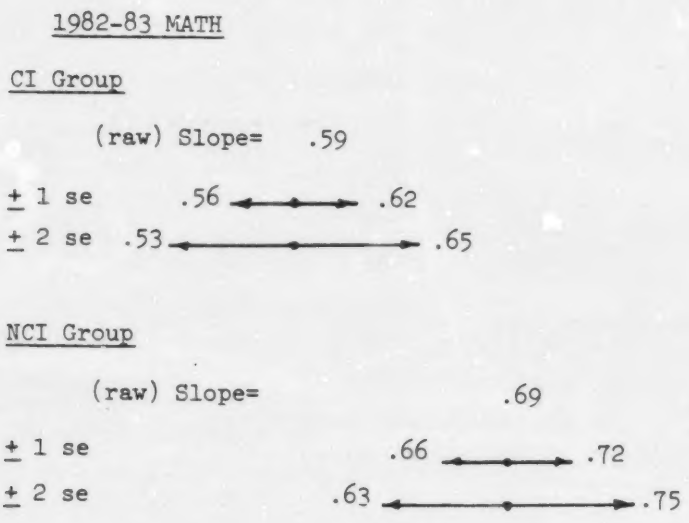
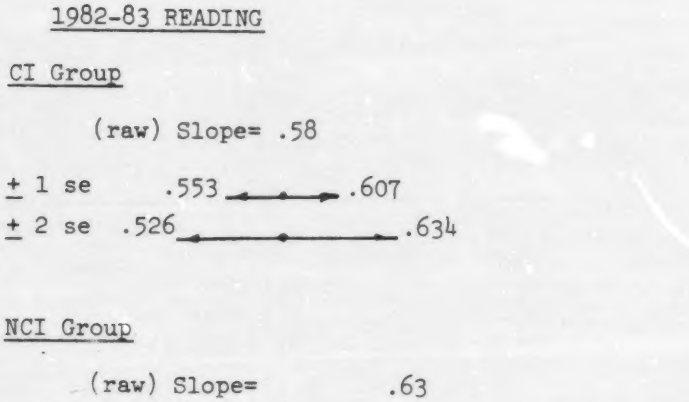



Figure 4

| 1982-83 MATH | 1982-83 READING |
|---|---|
| CI Group<br><br>(raw) Slope= .59<br>Intercept= 16.45<br>se Slope= .03 | CI Group<br><br>(raw) Slope= .58<br>Intercept= 14.89<br>se Slope= .027 |
| NCI Group<br><br>(raw) Slope= .69<br>Intercept= 12.12<br>se Slope= .03 | NCI Group<br><br>(raw) Slope= .63<br>Intercept= 13.32<br>se Slope= .027 |



Discussion. From a scientific perspective, the importance of these results is that they indicate that the "treatment" can affect students differentially within a compensatory program -- in this case, initially lower achievers apparently benefit more from treatment than the initially higher achievers. Accordingly, a simple between-group t-test, or even an ANCOVA applied naively would not have detected these differences. Obviously, the "homogenity of slopes" test is crucial, and has to be attended to prior to utilizing ANCOVA techniques. This preliminary test is more than mere statistical prelude to ANCOVA -- it addresses important substantive questions as well.

The study also had major implications for policy decision-makers. In these times of dwindling incoming resources and increased program costs, the point of intersection of the treatment slopes provided administrators with relevant data to modify the selection rule for the program, so that only those students who were most likely to benefit from the program would be served. This district had historically used the 40th percentile (nationally) as the "selection rule" or cutoff in both reading and math across the district. Within each designated Chapter I site, any student scoring lower than the 40th percentile was eligible for inclusion into the program. This often resulted in long waiting lists at some schools, because the number of students that were eligible

exceeded the maximum allowable as defined by the program regulations. With lower pre-score cutoffs for eligibility, the program comes closer to serving all those who qualify.

In math, in the 1981-82 data, the point of intersection indicated that the students below the 30th percentile (nationally) showed the most achievement gain due to Chapter I participation. In the 1982-83 math data, the students below the 39th percentile* appeared to benefit the most from the program. The recommendation for modifying the selection rule for math essentially involved "splitting the difference" between the 1981-82 and 1982-83 -- hence, it was recommended that the 35th percentile be adopted as the new cutoff.

In reading, in the 1981-82 data, the results showed that students initially scoring below the 15th percentile were the ones most benefitting from the Chapter I services. For reading, in the results for 1982-83, the percentile cutoff below which the Chapter I regression line was higher than the non-Chapter I line was the 18th. Administrators chose a cutoff for inclusion in Chapter I reading at a "convenient" value (the 20th percentile) just above the cutoff points indicated by the data. The reason for choosing a

*Even though policy stated that the 40th percentile ws the cutoff, some students were served with prescores up to the 50th percentile, so there were students in the analyses beyond the 39th percentile.

slightly higher value than suggested by the results for reading was to have an easy-to-implement eligibility rule, and because reading is recognized as a priority skill, and decision-makers would rather err in the direction of including a few too many students than perhaps exclude some students who may benefit from the treatment.

What psychological or pedagogical phenomenon could account for the observed results? Recent information-processing theory (Sternberg, 1984) has turned attention to "knowledge-acquisition" components (as distinct from "performance" components and other, higher-level "metacomponents"). Our findings may be due to the fact that different learner types may engage their "knowledge-acquisition" components quite differently. For illustrative purposes consider an oversimplified example.

Learner type #1 is at the 7th percentile overall in reading. This student has serious problems with <u>most</u> sub-areas included as part of Total Reading. However, learner type #2 is at the 35th percentile nationally, and is basically an "average" student in all reading subtests except for one, perhaps phonics, which has a particularly low score (that brought the overall reading score down to 35). The type of treatment "needed" by these two types of students might be quite different.

Student type #1 needs individual attention. In order to maximize his or her acquisition of knowledge about reading, a highly structured environment works best. The small class size and presence of 2 adults with only 8-10 students suit this student's needs quite well. He or she will gain the most if placed in a Chapter I Learning Center.

Student type #2 scores around the 50th percentile in all reading sub-areas except for phonics. This student is basically functioning normally in the regular classroom. He or she mainly needs some extra help in one area -- phonics. The gains of this student may be maximized by leaving him or her in the regular classroom. For this student, the benefits of the small class size and individual attention in Chapter I do <u>not</u> outweigh the costs incurred by losing the continuity of regular classroom instruction.

Such an explanation would imply that certain educational effects sometimes "compete" (and, further, that the nature and outcome of that competition varies from one type of learner to another). Student type #1 responds so well to the high structure of the Chapter I Learning Center that the loss of continuity in regular classroom instruction is more than overcome. However, the type #2 student does not benefit enough from that (Learning Center) environment to fully compensate for the loss of what would have been gained from staying in the regular classroom.

Specifying this "cost benefit" conception of Chapter I effects in combination with regular classroom effects requires future research.

## REFERENCES

Campbell, D.T., and Stanley, J.C. Experimental and Quasi-Experimental Designs for Research, Chicago, Rand McNally, 1966.

Crawford, W.J. A study of instructional processes in Title I classes. Journal of Research and Evaluation of the Oklahoma City Public Schools, Vol. 14, No. 1, June, 1983.

Crawford, W.J., Patrick, A.H., and Kimball, G.H. Relationships between minutes of program treatment and achievement gains in Chapter I students (1984, In Preparation).

Cronbach, L.J. and Snow, R.E. Aptitudes and Instructional Methods New York, Irvington Publishers, Inc., 1977.

Denham, C., and Lieberman, A. (Eds.) Time to Learn, National Institute of Education, California Commission for Teacher Preparation and Licensing, May, 1980.

Echternacht, G. Title I Evaluation and Reporting System: Development of Evaluation Models. In New Directions for Testing and Measurement, Measurement Aspects of Title I Evaluations. G. Echternacht (Ed.) San Francisco, Josse-Bass Inc., 1980.

Johnson, P., and Neyman, J. Tests of certain linear hypotheses and their applications to some educational problems. Statistical Research Memoirs, 1936, 1, 57-63.

Kimball, G.H. and Crawford, W.J. Compensating for underspecified models in the analysis of compensatory program treatment effects. Paper presented at the Annual AERA National Conference in Montreal, Canada, April, 1983.

Murray, S., and Arter, J. Internal Validity of the ESEA Title I Evaluation Models. In New Directions for Testing and Measurement, Measurement Aspects of Title I Evaluations. G. Echternacht (Ed.). San Francisco, Josse-Bass Inc., 1980.

Tallmadge, G. and Wood, C. User's Guide: ESEA Title I Evaluation and Reporting System. Mountain View, CA., RMC Research Corporation, 1976.

Sternberg, R.J. What should intelligence tests test? Implications of a triarchic theory of intelligence for intelligence testing. Educational Researcher, 1984, 13, 5-15.